California Franchise Tax Board (FTB)
Enterprise Architecture Council

Electronic Data Exchange Architecture Definition

Version No. 1.0

April 11, 2008

Author:
Enterprise Architecture Council

Document Information

Document Source

This document is controlled through Document and Deliverable Management. To verify that this document is the latest version, contact Enterprise Architecture.

Revision History

Version No.	Date	Summary of Changes	Revision Marks

Table of Contents

List of Figures

1.0 Executive Summary and Charter

1.1 Overview

Electronic Data Exchange (EDE) provides a set of standards for structuring information electronically exchanged between and within FTB and other businesses, organizations, government entities and other groups. EDE focuses on the internal infrastructure required to manage electronic data exchanges. The management of electronic data exchanges includes:

- Scheduling and tracking of data sends and receipts.
- Data formats and standards supported.
- Infrastructure used to map interface data to internal data and components/services that facilitate that transition.
- Confirmation of receipt or receipt of confirmation.
- Identification and notification of internal data owners.
- Movement of data to internal repositories and that includes systems or applications, services or interfaces, message queues, or databases and other data repositories.

The EDE infrastructure will be adaptable to various mediums of transfer including electronic file transfers and data sent through tapes or other media. The Data Exchange Architecture Definition touches on cleansing, formatting, and transformation of data that comes in from 3rd party[1] partners. The same data captured in our scanning systems and Internet systems will use the same services for cleansing, formatting, and transforming data.

1.2 Scope

Our EDE facility is a centralized secure data exchange method which provides built-in quality checks, supports standard file formats, and a common user-friendly approach to exchanging data without communication barriers. It provides a common set of data exchange formats and methods the sender and receiver can select to meet their technical capabilities. EDE crosses multiple lines of business and is combined in one robust framework that functions with cross-divisional applications.

This EDE architecture definition document further defines the current and target states of the FTB's EDE Architecture, a gap analysis, and a strategy for implementation.

1.3 EDE High-level Requirements

The following table outlines the high-level requirements of EDE.

Figure 1.3-1: EDE – Integrated Requirements

Requirement	Description
Flexibility	EDE services will support multiple types of communication such as FTP, secure HTTP, and web services.
Security	The communication of data will be secure and meet required standards, such as FIPS 141.
Configurable	New communication configurations will be provided through configuration interfaces that do not require programming. The services will also be able to hold the interface requirements of the external party, such as interface and data contract details.
Alerts	EDE services will provide for alerts that communicate success or failures of data exchanges.

[1] Note that third party data may include restrictions on access or use of the data consistent with state and/or federal law and the agreement with the data source.

Requirement	Description
Interfaces	EDE will support automated routing of data to internal systems, services, queues, or databases containing all routing details and requirements.
Audit Trail	EDE services will maintain an audit trail of all activity.

1.4 Conceptual Model

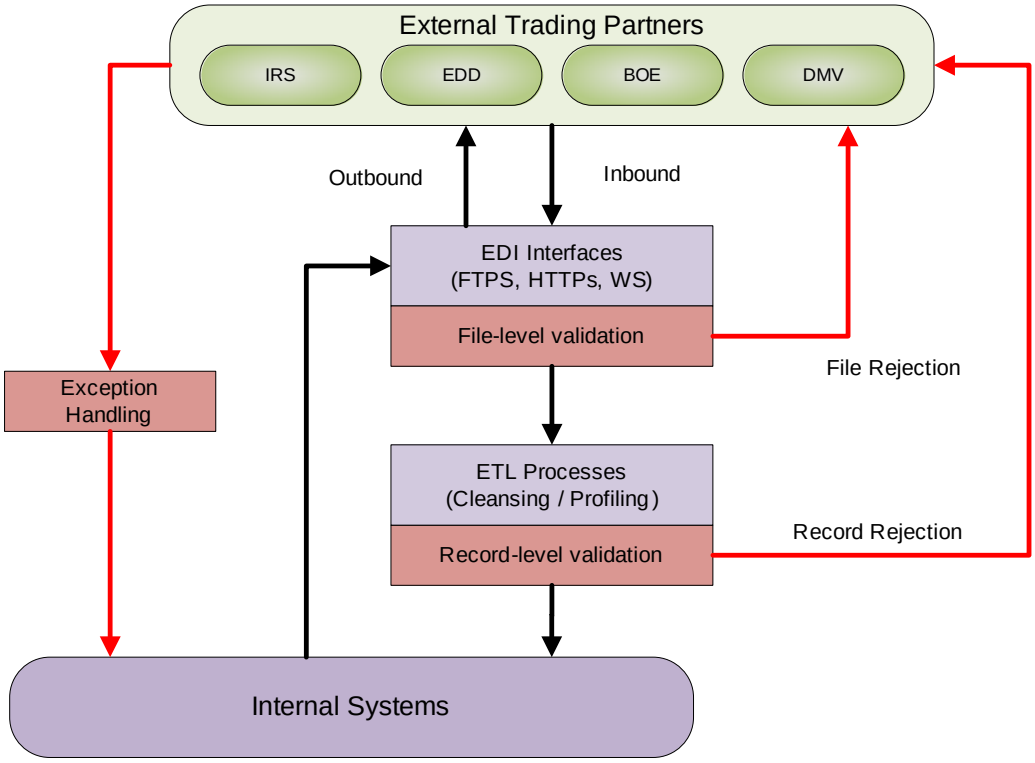
The following diagram shows services relevant to the EDE infrastructure at an enterprise level. The diagram below illustrates the processes involved with data exchange.

Data must be exchanged with external partners through data exchange interfaces (EDI). These interfaces will be open standards-based and support the most widely used protocols such as FTPS (Secure File Transfer Protocol) or HTTPs (Secure Web).

When files are received and checked for file-level corruption, the interface or data contract details must be stored and used. The file follows a standards-based file structure such as XML. Files that fail validation are returned to the sender for correction and resubmission, with possible information about the errors or omissions.

The ETL processes that cleanse and transform data for internal use is not in the scope of this architecture definition document.

Figure 1.4-1: ETA



FTB currently receives approximately 250 data sources[2] from direct reporters, federal agencies, California departments, and providers of purchased products. There are potential modifications to the existing systems that would be of benefit such as adding a central database for logging or

[2] Note that third party data may include restrictions on access or use of the data consistent with state and/or federal law and the agreement with the data source.

an SFTP protocol. Our current EDE receiving system is not yet positioned for tracking of record level information.

During the process of cleansing, transforming, and matching (see Data Delivery and Management Architecture Definition document), errors may be discovered that require records to be returned to the sender for correction and resubmission. Error reporting is a data exchange function.

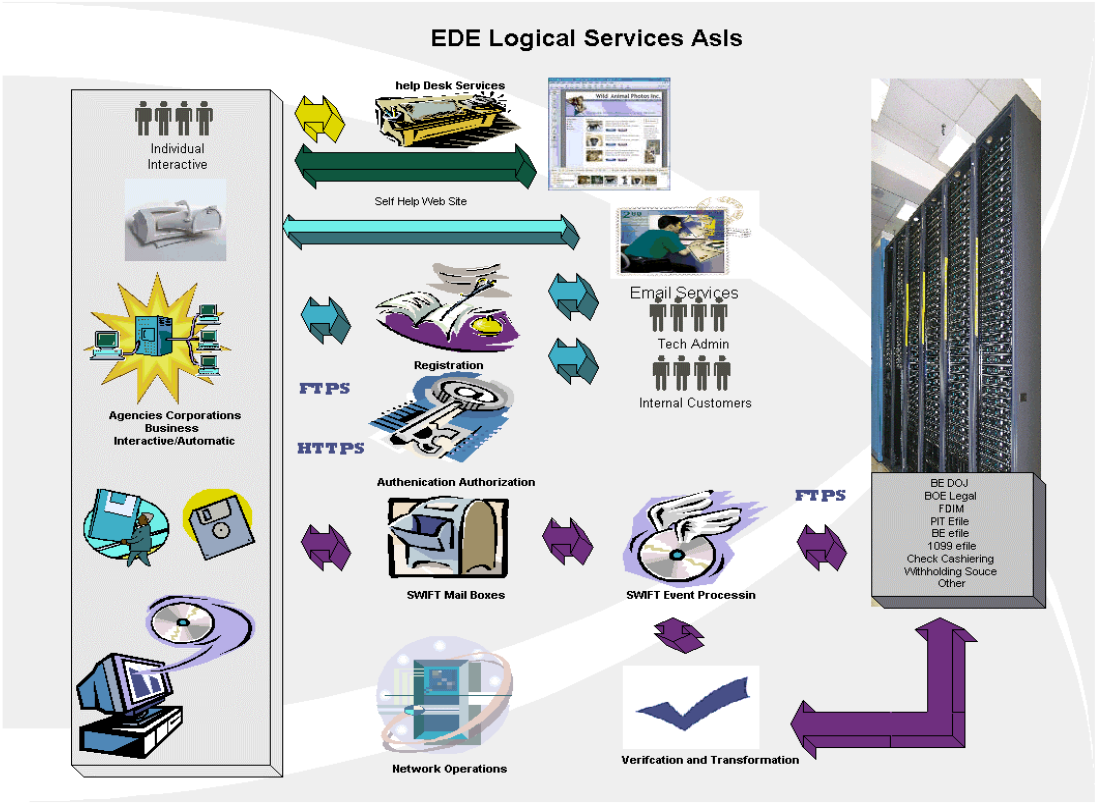
2.0 Current Architecture

The existing EDE System for FTB (Secure Web Internet File Transfer (SWIFT)) has these components:

- Website – general SWIFT portal.
- Registration – manual through already established business partner.
- Email services – SWIFT uses client to send email through an enterprise SMTP server.
- Transport service – through already established open standard protocols.
- Security authentication service – uses Active Directory. Initial entry is done manually. Two emails, one with user name and one with password are sent out to user unencrypted.
- Help desk services – through SOW or enterprise help desk.
- Monitoring services – logging within SWIFT tracks events.

Once the registration process is complete, the customer chooses FTPS or HTTPS standard protocols for transferring files. Customers can transfer or retrieve files waiting for them on SOW systems via SWIFT. Users use their own web browser or client that supports their chosen secure protocol. The system carries out a process when the customer logs onto the system and attempts to transfer or retrieve files depending upon which system they are transferring to or type of file they are transferring. If required, after the transfer is complete, SWIFT generates a receipt. If a customer transfers an e-file return, the system passes the file to the target system for verification. The receiving system returns an acknowledgement to SWIFT for the customer. SWIFT sends the acknowledgement, which contains the status of the transferred file.

Figure 1.4-1: Current EDE Logical Services



2.1 Initiation of Transfer

A common understanding of how data will be formatted is required to transfer data from one system to another. The data will be stored and transferred in logical predetermined data formats both systems understand. Exchange agreements, contracts, and IRS or FTB publications will be used to communicate the supported formats. There are fixed format files, comma delimited files, Extensible Markup Language (XML), Electronic Data Exchange Standard (EDI), Electronic Funds Transfer (EFT), custom data files, and unstructured data files. Unstructured files composed of many different types of information; such as, pictures and pdf files are included in data transfers.

A triggering action, either manual or automatic, is required to start the data interchange process. For an automatic data exchange, a timed process will be used, such as; a scheduler kicking off a program at certain intervals. Applications that are more complex might have a scheduling application built-in or be event driven. Manual initiation occurs when exporting and importing files and a person must physically start the process. Schedules that are difficult to determine in advance are usually candidates for manual initiation. Custom client, browser, or packaged file transfer software will be supported to perform manual actions. This mechanism enables the pushing and pulling of files. One can think of this as the electronic version of the U.S. post office.

2.2 Methods of Transfer

Method of transfer addresses both the protocol used and the type of interface. Customers are typically at different levels of technology expertise. Access to the EDE system will address a variety of customer technology. The number of protocols eligible for use by a customer may be limited by the customer's computer systems' technology or security. The protocol methods will be best practice protocols and Internet standards. Of the multiple ways to exchange data over the web, two standards, HTTPS and File Transfer Protocol Secured (FTPS), are the most common. Secure Shell (SSH) is a network protocol that allows data exchange over a secure channel between two computers. Secure File Transfer Protocol (SFTP) is SSH over FTP and is gaining in popularity with the rise of Unix/Linux platforms.

2.3 Transport

The EDE system will include the following for transport methods:

- Security, Security, and Security.
- Guaranteed delivery and restart.
- Proof of file integrity. Server will provide proof of data integrity of file transfers using Checksum or MD5.
- Event Driven API.
- Data privacy.
- Support for large file transfers.
- Support for batch files and interactive file transfers.
- Online help within browser based method.
- Support for all file types binary, text, etc.
- Eliminate redundant infrastructure and its associated costs.
- Testing and development systems for support.
- Regular security scans and assessment.
- Support for emerging protocols.
- Manage the file transfers.
- Email notification.
- Free client software without a license charge.

2.4 Data Transfer Mechanisms

The EDE system receives the file and then transfers the file to the destination using a standard protocol common with all FTB operating systems. The receiving system will be able to reply to the delivery system.

2.5 Provide user responses

Alerts, receipts, and acknowledgments are different methods for responding to a customer action or FTB action. Alerts can be email notices for error process or events. Receipts are the first response, if required, to denote that the file was received. Acknowledgements contain information about whether the file passed or was rejected. If the file contained errors, there is an explanation of the errors. The Testware tool is an offline method to provide user responses.

3.0 Target Architecture

3.1 Future Capabilities and Components

The future capabilities and components of our EDE system are the same as the current with slight modification:

- Registration – automated through a web self-registration system (See AIM Architecture Definition document).
- Security authentication service – uses Enterprise AIM system.
- Help desk services – through enterprise help desk.

EDE delivery methods are data agnostic providing business and government agencies secure Internet data transfer methods. The delivery methods utilize best practice protocols that allow system users to transfer both structured and unstructured data files. The security and self-registration are dependent on the Security Architecture Definition defined in the Identity and Access Management Architecture Definition document.

3.2 Future Enterprise Governance

A central team will be responsible for defining, publishing dimensions, and supporting the system. Our multiple lines of business will be combined in one robust framework. The team will have the following responsibilities:

- Provide cross-divisional support for applications.
- Identify implementation issues and work with appropriate stakeholders to resolve.
- Facilitate resolution of ambiguities within standards and rules.
- Coordinate with outreach to other related industry entities, identifying information gaps, and identifying strategies to fill those gaps.
- Establish consistent procedures for file transfers.
- Improve the quality, affordability, and availability of data to support tax business through effective and efficient information exchange and management.
- Assure the system complies with Sarbanes-Oxley (SOX), Gramm-Leach Bliley Act (GLBA), Health Insurance Portability and Accountability Act (HIPAA), and Federal Information Security Management Act (FISMA).

4.0 Gap Analysis

4.1 Gap Analysis Defined

The chart below contains the risks, benefits and critical success factors for the gaps between the existing EDE and the target EDE.

Figure 4.1-1: Gap Analysis Summary Table

GAP	Success Factor	Benefit	Risk/Issue
Unable to push/pull files from customer systems.	Able to push data files to customers and obtain files from them without human intervention.	Cost effectiveness and customer convenience.	Requires standardization of port ranges.
Manual self-registration.	Customers are able to register themselves for file downloads.	Cost effectiveness and customer convenience.	
Cost of Licenses is too high.	Replace software component where TCO is high. Contract renegotiated to lower price point.	Cost savings.	May cause costumers to relearn how to use system.

5.0 Roadmap

5.1 Phase Recommendation

Figure 5.1-1: Phase 1 Recommendation

Attribute	Business Opportunity/Requirement	Description of Effort
Self Registration.		Currently underway in the EASE project.
License cost.		Replace software component where TCO is high. Contract renegotiated to lower price point.

6.0 Appendix

6.1 Definitions

- **Reporting.** The reporting EDE function includes reports on transfers, required security logs, system logs, monitoring systems, and communication systems. The communication system will enable FTB to alert customers or internal staff regarding the status of transfers.

- **Registration and Self-Help Web**. This will allow a transmitter the ability to register for access to the EDE and manage their security accounts for authentication and authorization. The transmitter will be able to obtain self-help like resetting passwords, common ask question, and troubleshooting problems through a web interface.

- **Help Desk Services.** This complex service supports customers transferring files. The range of help can be from password resets to assistance understanding rejected files.

- **Testware.** Testware is downloadable software that will allow a transmitter to pre-check their file integrity before sending it to FTB. It will perform common field validation and format checks with an instant response. It will also contain converter software to convert from the transmitters file structure to an acceptable file structure. Testware is currently a stand-a-lone software product with limitations.

- **Fixed Format.** Fixed Format is a file structure consisting of physical records of a constant size within which the precise location of each variable is based on the column. Programs receiving fixed format data would be programmed to receive data in that exact format. This is typically documented by a publication produced by FTB.

- **CSV (Comma-Separated Value).** CSV file format is a common text file format that contains comma-delimited values. CSV is almost universally supported by applications, but it poses challenges as well. The data cannot contain quotes and commas due to those being used for field delimitation. As with fixed formats, the receiving system will still need to be programmed (or mapped) to know what type of data it is receiving.

- **XML.** The XML format is known as a self-describing format, which uses field tag names to delineate the values. The information about the data, such as field names and types, is encoded with the data, so a receiving system can dynamically receive it and dynamically map the data to the database, making the data transfer process less laborious than with a CSV or fixed format file. XML is the standard for transferring data over the Internet although XML requires additional computing power to process the records.

- **Electronic Data Interchange Standard.** The Electronic Data Interchange Standards is a set of standards for structuring information electronically exchanged between and within businesses, organizations, government entities, and other groups. EDI is still the data format used by the vast majority of electronic commerce transactions in the world. EDI documents generally contain the same information that would normally be found in a paper document. EDI is used among many professions, legal, transportation, shipping, etc. EDI has not found its way into tax form processing.

- **Electronic Funds Transfer (EFT).** Financial institutions where money is transferred from one account to another utilize EFT. Examples of EFTs include; electronic wire transfers, automatic teller machine (ATM) transactions, direct deposit, business-to-business payments, web payments, and federal, state, and local tax payments.

- **Custom Data Formats**. Custom data formats do not meet newer standards, however, may be driven from the originators of the data.

- **Unstructured Data Files.** Unstructured data files are a group of files composed of many different types of files such as; documents, images, multimedia, and spreadsheets.

- **HTTPS (Hypertext Transfer Protocol Secure)** HTTPS provides for secure transmission and receipt of data the majority of which occurs via the Uniform Resource Identifier (URI). An HTTPS request can be typed or linked into the address bar of a web browser, or coded within an application. HTTPS request types are commonly called web services and include Representational State Transfer (REST) and Standard Object Access Protocol (SOAP).

- **REST** stores the transferred information in a packet that is sent along with the URI using a standard POST transaction within HTTPS. This can be initiated from a browser or a custom application. The technology to use POST is available on any Internet connected system. HTTP is the standard technology.

- **SOAP** transfers data packets via the POST function requiring specific format rules in order for the data packets to be properly read, and can include additional instructions or procedures.

  HTTPS can use REST to transfer data directly via URLs and POST parameters, whereas SOAP transfers all data in packets via the POST function. This makes REST an easier method to use, and SOAP a more powerful method. SOAP based transfers are usually initiated from a software program.

- **File Transfer Protocol Secured (FTPS)** FTPS allows you to send data from one place to another using FTP through a Transport Layer Security (TLS) and its predecessor, Secure Sockets Layer (SSL), connection.

  - Generic Client – there are many clients available to allow human interaction to transfer files and pick up files (ex. Microsoft Internet Explorer, FTP Pro, Cute FTP, Fire Fox, etc.).

  - Custom client – is the same as a generic client, however, may offer schedule option to eliminate human intervention (ex. FTP Pro, Tumbleweed Secure Transport Client, etc.).

  - Program – the interaction is completed by computer-to-computer connections. The programs use FTPS as a protocol and transfers data based upon the workflow of vendor systems. Usually the most technically sophisticated users have complex fully automated systems.

- **SFTP** SFTP is a network protocol that provides file transfer and manipulation functionality over any reliable data stream. This is an emerging protocol, which is not yet an Internet standard.

  - Generic client – there are many SSH products that can allow a user to interact with a command line and GUI interface to transfer a file. Most Unix systems have a built in Secure Shell to access the computer. There are other products such as Putty, Cygwin, etc.
  - Program – the interaction is completed by computer-to-computer connections using SFTP protocol to connect and transfer files.

- **Other Media Formats**

  Not all customers have the technology to transfer data over the Internet. Many customers still prefer using magnetic media, such as disk or tape, and transfer the data using postal services. Magnetic files are uploaded on behalf of the customer and processed through the system. The acknowledgement is processed via postal service or email. The manual processing of these files will be incorporated within the EDE system and reuse the same programs for processing the data.